electronics

Article

Suicide Bomb Attack Identification and Analytics through Data Mining Techniques

Faria Ferooz [1,*], Malik Tahir Hassan [1], Mazhar Javed Awan [1], Haitham Nobanee [2,3,4,*], Maryam Kamal [1], Awais Yasin [5] and Azlan Mohd Zain [6]

1 Department of Software Engineering, University of Management and Technology, Lahore 54770, Pakistan; tahir.hassan@umt.edu.pk (M.T.H.); mazhar.awan@umt.edu.pk (M.J.A.); maryam.kamal@umt.edu.pk (M.K.)
2 College of Business, Abu Dhabi University, Abu Dhabi 59911, United Arab Emirates
3 Oxford Centre for Islamic Studies, University of Oxford, Marston Rd, Headington, Oxford OX3 0EE, UK
4 Faculty of Humanities & Social Sciences, University of Liverpool, 12 Abercromby Square, Liverpool L69 7WZ, UK
5 Department of Computer Engineering, National University of Technology, Islamabad 44000, Pakistan; awaisyasin@nutech.edu.pk
6 UTM Big Data Centre, School of Computing, Universiti Teknologi Malaysia, Skudai Johor 81310, Malaysia; azlanmz@utm.my
* Correspondence: faria.ferooz@umt.edu.pk (F.F.); nobanee@gmail.com (H.N.)

Abstract: Suicide bomb attacks are a high priority concern nowadays for every country in the world. They are a massively destructive criminal activity known as terrorism where one explodes a bomb attached to himself or herself, usually in a public place, taking the lives of many. Terrorist activity in different regions of the world depends and varies according to geopolitical situations and significant regional factors. There has been no significant work performed previously by utilizing the Pakistani suicide attack dataset and no data mining-based solutions have been given related to suicide attacks. This paper aims to contribute to the counterterrorism initiative for the safety of this world against suicide bomb attacks by extracting hidden patterns from suicidal bombing attack data. In order to analyze the psychology of suicide bombers and find a correlation between suicide attacks and the prediction of the next possible venue for terrorist activities, visualization analysis is performed and data mining techniques of classification, clustering and association rule mining are incorporated. For classification, Naïve Bayes, ID3 and J48 algorithms are applied on distinctive selected attributes. The results exhibited by classification show high accuracy against all three algorithms applied, i.e., 73.2%, 73.8% and 75.4%. We adapt the *K*-means algorithm to perform clustering and, consequently, the risk of blast intensity is identified in a particular location. Frequent patterns are also obtained through the Apriori algorithm for the association rule to extract the factors involved in suicide attacks.

Keywords: counter terrorism; clustering; geopolitical situation; location sensitivity prediction; pattern extraction; suicide; bombing; environment; data mining; big data

Citation: Ferooz, F.; Hassan, M.T.; Awan, M.J.; Nobanee, H.; Kamal, M.; Yasin, A.; Zain, A.M. Suicide Bomb Attack Identification and Analytics through Data Mining Techniques. *Electronics* **2021**, *10*, 2398. https://doi.org/10.3390/electronics10192398

Academic Editor: Amir Mosavi

Received: 25 August 2021
Accepted: 28 September 2021
Published: 30 September 2021

1. Introduction

Suicide is a global issue in the world nowadays. There are many factors that are affecting frequently escalated suicide bomb attacks such as unsatisfactory leadership, inequality, religious conflicts, weak judicial systems and a lack of basic necessities for people, causing psychological disturbances. Suicide bombings are an enormously worthwhile approach of terrorism. Suicide attackers have become one among the foremost feared terrorist weapons [1–3]. Suicide attackers mainly carry explosive charges on their bodies or carry such charges in different vehicles, such as a car, truck, ship or an airplane. Suicide bombing is described as "a politically inspired violent attack carried through an individual who is completely aware and deliberately motives his or her own death as well as other people". The suicide attacker's death is the pre-condition for the success of the blast and their aim is to build fear among people [4,5].

Suicide terrorism is frequently examined either from a security and policy perspective usually through information retrieval techniques and information extraction [6,7]. Most terrorist groups utilize suicide operations at least in part since they are the most crowded. They have found out the most suitable ways to kill and injure the maximum number of people; in this manner, the strategy has multiplied broadly and has been used with higher frequency. In simple terms, terrorist groups are selecting the proficient strategies accessible to them to produce the desired consequence [8,9].

Recently, machine learning (ML) or data mining algorithms have been applied in many domains through supervised and unsupervised learning techniques [10–14].

The Figure 1 shows the mechanism of suicide bombers through data mining techniques.

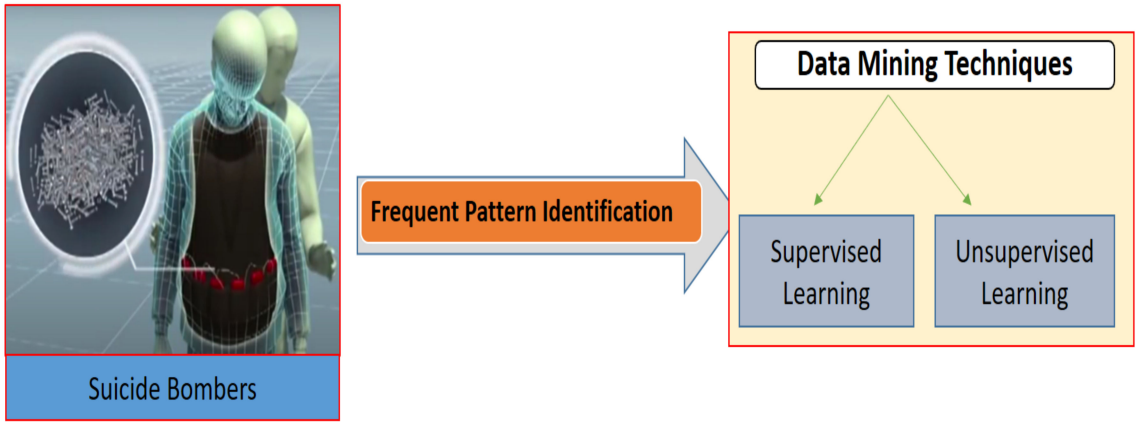

Figure 1. Suicide bombers' frequent pattern identification mechanism.

As a counter terrorism initiative, this study aims to state a prediction for the probable venue and location for upcoming suicide bomb attacks, based on previous recorded data and statistics for such events happening in the past. As the identification of and implications for the upcoming attack depend on a series of previous records of similar incidents, a strategy to group such events based on common features and categories is required.

Furthermore, we also aim to discover frequent patterns for suicide bomb attacks, which can be expected to happen at a particular location based on the strategies used in previous attacks.

Pakistan is one of the countries where terrorism is a frequent threat. It has faced numerous suicide bombing incidents with multiple casualties. There has been no major effort performed earlier by utilizing the Pakistan suicide attack data, and no significant data mining-based solution has been provided on suicide attacks. Therefore, due to concerns of civilians and the government, we have specified the problem domain and datasets for this state [15]. Our adopted techniques and datasets are confined to Pakistan for the purpose of this research. A detailed analysis is performed on the most recently updated dataset for suicide bombing attacks from 1995 to 2017 taken from Kaggle [16].

To cover more recent datasets, we collect the remaining data manually from various studies [17,18].

In this study, visualization analysis is performed and both supervised learning (i.e., classification) and unsupervised learning (i.e., clustering and association rule) techniques are leveraged. In the supervised learning model, multiple classification algorithms, Naïve Bayes, ID3 and J48, are used to predict the next possible venue for a suicide attack, whereas in unsupervised learning the label of the classes is unknown. Thus, a clustering algorithm, i.e., K-means, is used to group the related suicide attack characteristics [19]. In addition, an association rule mining algorithm, i.e., Apriori, is used to observe and analyze the frequent patterns of suicide attacks and to identify the expected location.

The main contribution of this paper is to predict the suspected location for upcoming attacks by analyzing common features and categories for previous attacks. Apart from this,

it is worthwhile to identify characteristics for similar suicide attacks and suicide attack strategies and trends for expected targeted locations. This will not only help in the tracking of the records of past suicide bombs, but also help in the prediction of future attacks, thus making this world a secure place to live and saving the lives of millions.

The contributions of this paper are summarized as follows:

1. Our approach identified suicide bomb attacks pattern through data mining algorithm of extensive Pakistani datasets.
2. The data were further pre-processed to make them useful for knowledge extraction.
3. We performed data visualization to analyze and explore the major causes of suicide attacks.
4. We predicted the next possible location for a suicide attack using previous recorded data based on supervised data mining techniques.
5. Lastly, we identified similar suicide attacks and frequent patterns for upcoming and expected suicide bomb attacks based on unsupervised data mining techniques.

This paper is organized as follows. Section 2 provides the literature review. Section 3 illustrates the dataset and methodology. Section 4 shows the experimental results and discusses them. Section 5 describes the conclusion and future work.

2. Literature Review

In this section, research articles applying data mining techniques to crime datasets are referred to and discussed in detail. Moreover, the limitations of the studies are also discussed.

As the concerns of suicide bomb attacks and related sufferings are growing with each passing day, researchers, government officials and armed forces in every country are working in collaboration to eliminate these threats. There have been multiple appreciable research efforts to analyze suicide bomb attacks and identify patterns or co-relations. They performed multi-level analysis techniques of suicidal bomb attackers in Pakistan, including their personal, marital, demographic and economic traits on the basis of data gathered from primary and secondary research efforts, compared Pakistani suicidal bomb attackers with attackers from other countries (e.g., Turkish, Palestinian and Lebanese attackers) and performed classification based on ages and educational aspects [20].

Asal et al. [21] analyzed 14 hypotheses gathered from social, political sciences and organizational literatures to deduce collaborative patterns among attackers. They presented a motivation and region-based clustering among the attackers and interest-based cooperation.

Filote et al. [22] stated that such attacks are related to religious conflicts. They showed that religious conflicts were the major cause of suicide attacks from 1981 to 2010. The study identified the suicidal attacks in a particular area and its detailed information can predict terrorist activities in that particular area.

Cozza and Rubino [23] introduced an approach that groups similar terrorist suicidal attacks in terms of time, space, dimensions and keywords into clusters. They organized and mined blocks of information used to measure security measures.

Malik et al. [24] identified major risk factors of such attacks. They indicated that there are 65 risk factors that are involved in terrorism, 13 of which are significantly crucial. Classification based on leadership dishonesty and disloyalty towards a country such as a lack of basic necessities, corruption, inequality and poor judicial systems was also conducted.

Verma et al. [25] examined research to predict the identification of the perpetrator of a terrorist incident using machine learning techniques on historical data and showed effective working for data analysis.

Diab [26] worked on the automated classification of terrorist attacks globally and enhanced Stochastic Gradient Descent (SGD) algorithm for this purpose. They examined different settings for representation, transformation and weighting features from the descriptive summary of recorded terrorist attack incidents through the Global Terrorism Database as a pre-classification step and conducted SGD learning on support vector machine (SVM), logistic regression and perceptron classifiers. They showed considerable accuracy and execution time for analysis.

Saeed et al. [27] identified similar patterns from bombing attack data and concluded that patterns and methods of terrorist activities, especially suicide bombing, varied according to geopolitical situation. They showed that terrorist incidents in Khyber Pakhtunkhwa (KPK) during the last three decades have varied from 5% to 45% and have grown to 52%, while in Sindh the pattern is opposite—terrorist activities have grown from 23% to 68% and have moved down to a range of 9%. After a detailed analysis of terrorist activities, they identified that terrorist activities varied according to the political and social situation of a region.

Imana and Kirk [28] proposed a technique showing that row-wise crowd formation is the best pattern to reduce the number of fatalities to 12% and injuries 7%. The results came out from simulations compared to real-life incidents and showed it is good enough to show the impacts on counterterrorism.

Conlon et al. [29] analyzed information regarding intelligence and security issues available in documents in printed and electronic format. They proposed a technique based on the knowledge of engineering, named CAINES, to extract information such as intelligence, attacks and security from electronic documents.

Saiya and Scime [30] suggested religious aspects and related factors as the common cause of terrorism attacks. The data mining approach clustering is applied of the religious terrorism dataset to find significantly contributing attributes.

Tutun et al. [31] identified unexpected interactions through using non-similarities among attacks. The approach was used to find the possible outlier by analyzing the past strategies used in the events. The results showed that, by comparing the events and the detected patterns, it can match with more than 90 percent in terms of accuracy.

Uddin et al. [32] explored artificial intelligence and machine learning techniques to understand the behavior of terrorist activities. Five different models based on a deep neural network (DNN) were created to understand the behavior of terrorist activities. The designed models used a single-layer neural network (NN), five-layer DNN and three traditional machine learning algorithms, i.e., logistic regression, SVM and Naïve Bayes.

JSPM and Tirwa [33] worked on predictive modeling for terror attacks where three predictive models were designed to classify attack types and attack regions and weapon types based on millions of attacks using various supervised machine learning algorithms. The models used support vector machine, 49 Forest (RF), REP Tree and J48 for the classification.

Huamaní et al. [34] used AI techniques to quantify and predict possible terrorist attacks using decision trees and random forest classification models. However the study did not identify patterns of terrorist attacks.

Kirk [35] noted that suicide bombing has become one of the most popular modes of operation for terrorist organizations all over the world. In order to access the impact of crowd density on suicide attackers' benefit, a simulation tool was developed to analyze the specifics of crowd formation and bomber orientation with respect to crowds. The objective was to analyze optimal crowd formation to reduce crowd deaths and/or injuries. The findings can be utilized to make a plan for post-disaster management and counter-terrorism.

Tayal et al. [36] used a variety of data mining techniques to detect crime in Indian crime data. The unstructured data were collected from various web sources between 2000 and 2012. The analysis was carried out using clustering and classification algorithms. For clustering, the K-means algorithm was used to group similar criminal activities. For classification, K-Nearest Neighbor was used to predict crimes that are likely to occur in the future. They reported that its accuracy was 93.2 percent.

The research and efforts to identify, classify and predict terrorism attacks are still in progress and researchers from all over the world try their best to cater to these concerns. Our proposed methods and research involving classification, clustering and attack pattern recognition are discussed in Section 3.

3. Materials and Methods

3.1. Data Collection and Preprocessing

To show the effectiveness of the proposed methods according to the research objectives, we used a dataset that specifically recorded suicide bomb attacks that occurred in Pakistan. The total collected dataset consists of a rich categorical data deposit with 514 instances and 24 attributes (i.e., Date, Islamic Date, Blast Day, Holiday Type, Time, City, Latitude, Longitude, Province, Location, Location Category, Location Sensitivity, Open/Closed, Influencing Event, Target Type, Target Sect, Killed Min, Killed Max, Injured Min, Injured Max, No of Suicide Blasts, Explosive Weight, Temperature (C) and Temperature (F)). The dataset was collected from two sources.

First, the dataset was taken from 1995 to 2016 with 496 instances from Kaggle [16]. Second, we collected the rest of the recorded data onward from 2017 from different web sources and recorded them manually against all the attributes in the suicide attack dataset [17,18]. As a result, a total of 18 recorded suicide attack details were added in the dataset.

We pre-processed the dataset using the NLTK package, which is a Python library, to extract desired information from raw datasets according to the requirement of our research purpose [37].

In the pre-processing phase, we smoothed out the data by removing data noise and errors, which cause a hindrance to loading these data on WEKA in ARFF format. WEKA is a collection of machine learning algorithms frequently used to perform data mining tasks. It has a rich collection of strategies and algorithms for data pre-processing, classification, regression, clustering, association rules and visualization [38].

During this process, we found some missing values in some attributes such as Target Type and Location Sensitivity. We filled the missing values by the respective mean and mode of each of the particular columns. For better analysis, data discretization was performed on the selected attributes. The attribute Blast Day was derived from the Date attribute. We further simplified the attributes Location Category (Targeted Location) and Killed Max (Number of People Killed). Finally, we selected 6 attributes from this dataset below along with their possible values.

Blast Day (Holiday, Working day, Weekend)
City (Names of cities in Pakistan)
Location Category (Commercial Civilian, Educational Institutes, FATA, Foreign Embassy, Government Building, Hospital, Hotel, Market, Military Check-post, Office Building, Park-Ground, Police Checkpoint, Religious Places, Transport Area)
Target Type (Foreigner, Religious, Military, Government Official, Civilian, Police, Shia Sect, Anti Militants)
Location Sensitivity (High, Medium, Low)
Number of People Killed (Maximum number of people killed)

3.2. Data Visualization

To obtain insights from the pre-processed suicide dataset, we first performed data visualization. In order to find out the type of people targeted in suicide bombing, a graph was generated. As per the analysis shown in the Figure 2, it can be clearly seen that the ratio of target people in suicide attacks for 'Military', 'Police' and 'Civilian' is high, while that for 'Shia Sect' and 'Foreigner' is low.

The distribution of the number of people killed in suicide bombing is illustrated in the Figure 3. It is clearly shown that, in suicide blasts, the ratio of killed people in 'Civilian' and 'Military' categories is high, whereas that for 'Shia Sect' and 'Foreigner' is low.

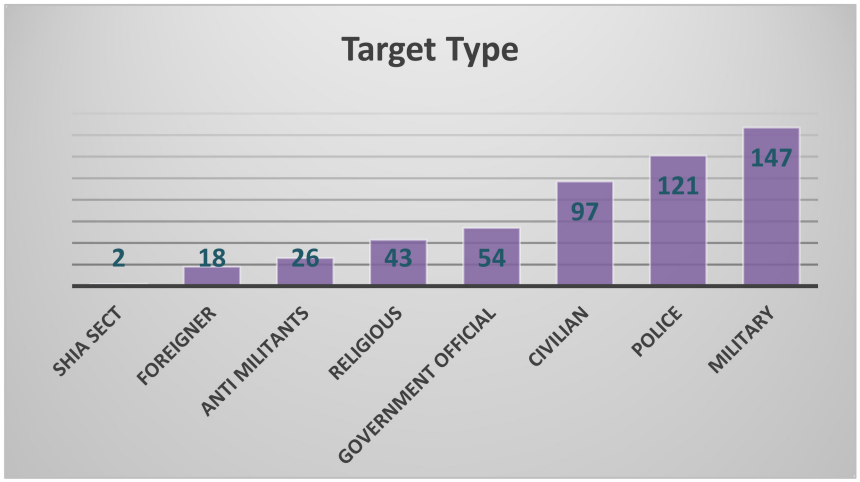

Figure 2. Visual analysis of the type of people targeted in suicide bombing.

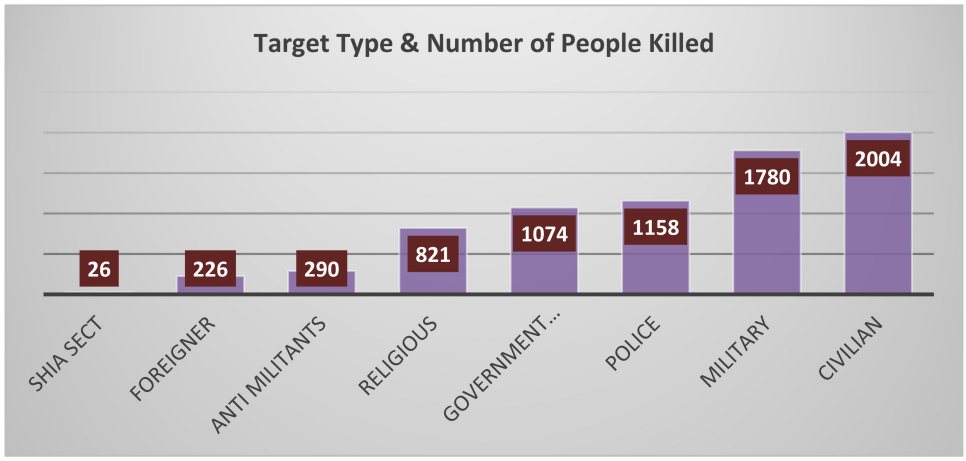

Figure 3. Visual analysis of the distribution of target type and ratio of people killed.

Further analysis is made on suicide attack locations, which are frequently targeted by suicide bombers; a graph has been generated, as shown in the Figure 4. As it is clearly shown, suicide bombers mostly target 'Police checkpoint' and 'Market' rather than 'FATA' and 'Banks'.

Figure 4. Visual analysis of the locations targeted by suicide bombers.

Further, in order to find out the sensitivity of locations for suicide attacks, we estimated a Radar chart, which indicates that the highly sensitive location is 'Police checkpoint' compared to medium- and low-sensitivity locations such as 'Religious palaces' and 'Park-Ground'.

Here, we can infer that suicide bombers mostly attack highly sensitive locations, as shown in the Figure 5.

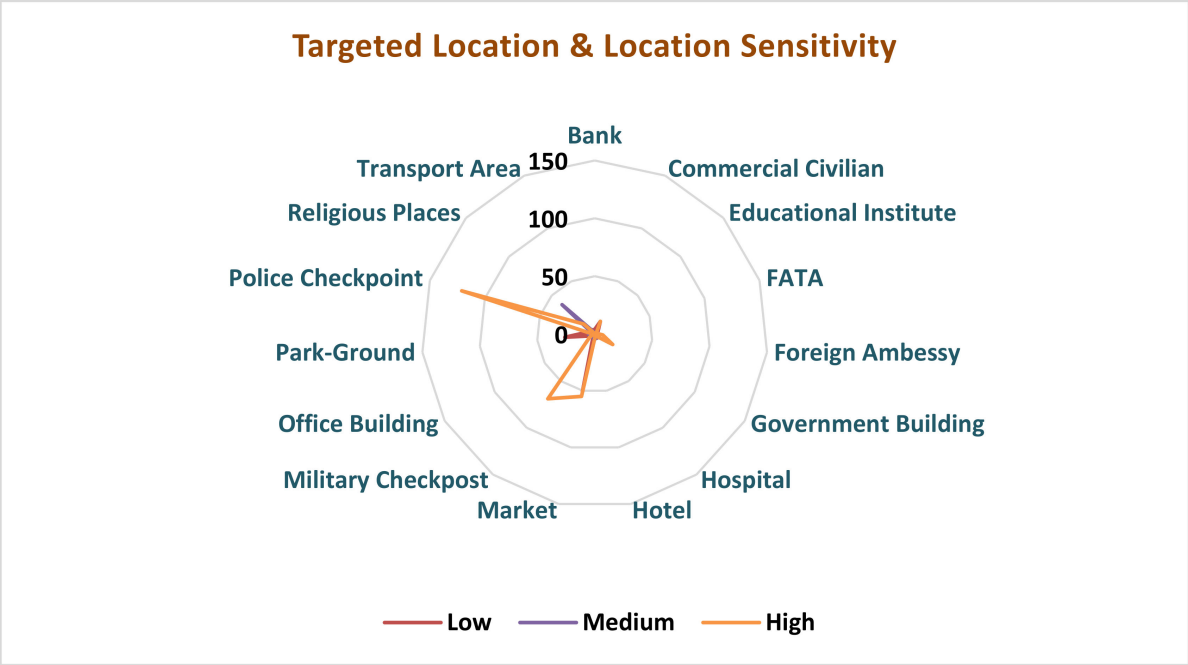

Figure 5. Visual analysis of the distribution of the targeted location by suicide attackers and sensitivity of that particular location.

3.3. Our Approach

In order to achieve the research objectives, various data mining techniques of supervised and unsupervised learning are applied on the suicide attack dataset. The Figure 6 shows the overall process of the proposed method. First, we gather the suicide bombing attack dataset. Second, pre-processing techniques are applied on the dataset. Finally, (1) visualization analysis is performed to identify critical findings from the dataset and multiple data mining techniques are applied. Specifically, (2) data mining classification techniques including Naïve Bayes, ID3 and J48 are applied to predict the occurrences of suicide attacks in particular areas. (3) Data clustering based on *K*-means is applied to identify similar characteristics of suicide attacks. (4) An association rule mining algorithm, Apriori, is used to extract the factors involved in suicide attacks.

The details of each process are discussed in the following subsections.

3.4. Classification Techniques

Classification is the most commonly problem in the field of machine learning. The supervised learning method, in which a set of pre-defined examples is used to construct a model that can classify instances of attributes at a large scale. Within the context of the given dataset, classification is used to generate several models of unidentified patterns and to assess prospects based on previous decision making.

We applied classification on the considered pre-processed dataset to classify and identify patterns of attacks based on statistics. We used three classification strategies: (1) Naïve Bayes, (2) Decision Tree ID3 and Decision Tree J48.

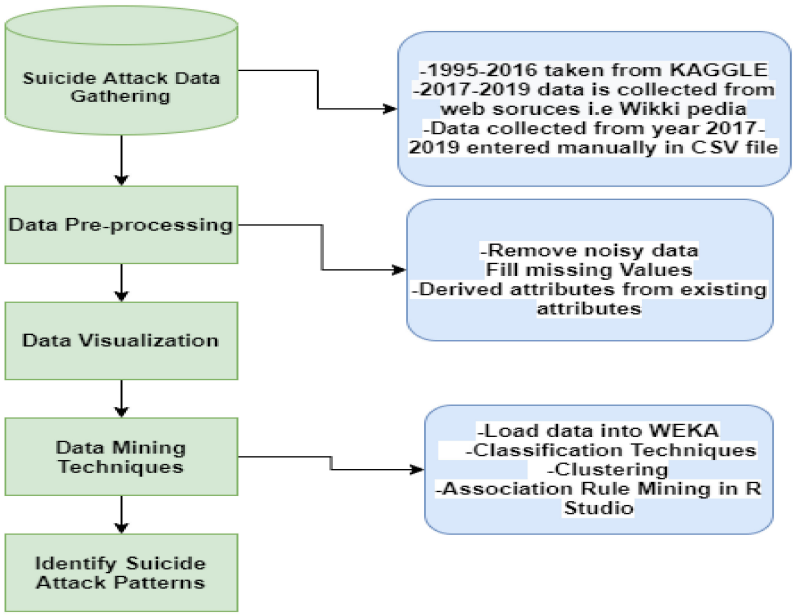

Figure 6. The overall process of the proposed method for suicide attack analysis and prediction.

3.4.1. Naïve Bayes Classifier

Naïve Bayes is a widely used classification algorithm for determining the accuracy of a dataset. For multi-class prediction problems, Naïve Bayes is a good choice. The Naïve Bayes classifier is based on the Bayes rule of conditional probability. It utilizes all the instances contained in the data and investigates them individually as though they are likewise important and independent of each other. It was introduced into the text retrieval and information extraction community early in the 1960s and it is still a common technique for text classification and the problem of judging documents [39].

The number of parameters required by the Naïve Bayes classifier is linear in the number of variables (features/predictors) in a learning problem. Instead of costly iterative approximation, which is used for many other forms of classifiers, maximum likelihood training can be obtained by evaluating the closed form expression, which takes linear time. Naïve Bayes models are referred to by many names. Simple and free Bayes are two terms used to describe Naïve Bayes models. These names refer to the use of the Bayes theorem in the classifier decision rule.

3.4.2. Iterative Dichotomiser 3 (ID3)

ID3 stands for Iterative Dichotomiser 3 because this algorithm iteratively divides its features into two or more groups in each step. It generally uses a top-down approach to build a decision tree. It builds a tree based on the information obtained from the training instances and then uses the same case to classify the data. Generally, it works well for attributes with no missing values [40].

It is one of data mining's predictive modeling techniques. In order to build the tree, the entropy measure is used to determine the nodes. Since greater entropy attributes create more uncertainty in the result, they are chosen in the order of the entropy. By dividing the source set into subsets on the basis of an attribute value test, a tree can be learned. This process is recursively performed on each derived subset, which is known as recursive partitioning. When the subset at a node has all the similar values of the target variable or when splitting no longer adds values to the predictions, the recursion is finished. This is the well-known approach for learning decision trees from data and is based on the top-down formation of decision trees.

3.4.3. Decision Tree J48

The J48 algorithm is also one of the best machine learning algorithms for categorizing and continuously examining data. The Decision Tree Algorithm discovers how the attributes act for various instances. Likewise, on the basis of the training instances, the classes for recently produced instances are found. This algorithm creates the rules for the prediction of the target variable. The critical distribution of the data is easily understandable with the assistance of the tree classification algorithm. J48 is an extension of ID3. The extra features of J48 account for missing values, decision tree pruning, continuous attribute value ranges, the derivation of rules, etc. [41].

The J48 algorithm builds the tree using the pruning technique. Pruning reduces the size of the tree by eliminating relevant data, which guides the terrible concert in prediction. J48 classifies data until the categorization is completed, providing maximum accuracy over training data. It provides classifier results in the form of rule sets and a decision tree. The results of each classification technique are presented in the Results and Discussion section.

3.5. Cluster Analysis

As mentioned earlier, clustering refers to the strategy of grouping together similar objects based on their common attributes. It is an unsupervised learning in which the class label is unknown. Thus, clustering-based models specifically help in the identification of suicide attack patterns. To find out the similar features and characteristics in suicide attacks, clustering plays an important role. We consider the K-means algorithm for clustering to help in the identification of suicide attack-related patterns. The K-means algorithm is applied on the suicide attack dataset using the WEKA tool to group similar types of suicide attacks that happened in a particular location.

K-means is one of the most widely used data mining clustering techniques. It is applied to the dataset. The goal of the *K*-means algorithm is to find groups from data and to group a number of groups represented by label K. K-means splits the data into groups whose members belong together and each object is assigned to its more related group. It is the most common clustering algorithm that groups data with similar features or characteristics together. The groups of data are called clusters. Each cluster group is similar between themselves but dissimilar from others [42].

K-means divides the data into clusters or groups with the following steps:

1. The target number of clusters, K, is given along with the input datasets.
2. A set of K instances is selected as the centers of the clusters.
3. The process considers each instance and allocates it to its nearest cluster.
4. K-means cluster centroids are recalculated after each instance assignment or after each cycle of re-assignment.
5. The steps from 2 to 4 are repeated until K clusters are obtained.

The result of the *K*-means algorithm is described in the Results and Discussion section.

3.6. Association Rule Mining

Association rule mining is a kind of unsupervised data mining technique to find out the frequent and interesting patterns in the dataset. The Apriori algorithm follows the association rule mining approach for the extraction of frequent *K*-item sets. It utilizes the downward closure property, which demonstrates that if any *K*-item sets are frequent, all of its subsets must also be frequent as well [43].

In this study, the Apriori algorithm is applied with the R language to find frequent patterns of suicide attack activities. In order to measure the association of the frequent patterns mined, three values, i.e., support, lift and confidence values, of each rule are calculated for each corresponding rule ($X \rightarrow Y$) [44].

For a rule ($X \rightarrow Y$), the support rule is labeled as sup ($X \rightarrow Y$). The value of support indicates the frequent occurrences of X and Y in the dataset together. According to Equation (1), the support value is defined as the frequency of the number of transaction ($X \rightarrow Y$) appearing in the database divided by the total number of transactions *N*. We can determine

that an item is considered frequent if its support value is equal to or greater than a specified minimum support value. The result is the frequent *K*-item sets.

$$Support = \frac{Frequency(X, Y)}{N} \quad (1)$$

Confidence specifies the correctness of the rule. The confidence is defined as the frequency of the number of transactions where ($X \rightarrow Y$) occurs together divided by the number of transactions where X appears. Confidence is calculated according to Equation (2). The greater the confidence value mean the association rule is stronger. If various rules have the same confidence value, they are organized on the basis of their support values.

$$Confidence = \frac{Frequency(X, Y)}{X} \quad (2)$$

To find the interestingness of an association rule, the lift measure is used. Lift is computed on the basis of support values collectively as well as individually of each frequent item in the item set, as shown in Equation (3). That is, support is divided by the multiplication value of $Support(X)$ and $Support(Y)$, which are the number of transactions containing X and Y, respectively [44]. The higher value of the lift indicates the interestingness of that rule, which means the frequent occurrence of X and Y together.

$$Lift = \frac{Support}{Support(X) * Support(Y)} \quad (3)$$

The locations are analyzed at risk of attacks based on the previous such activities, we apply the Apriori algorithm to our suicide attack dataset for identifying frequent suicide attack patterns.

The purpose of identifying frequent patterns is to analyze suicide attack activities, which mostly occur at a particular location (i.e., Location Category in the dataset) and sensitivity level of that location (i.e., Location Sensitivity in the dataset). Frequent pattern analysis also helps in the verification of the results obtained from the clustering solutions.

The results of the Apriori algorithm are described in the Results and Discussion section.

4. Results and Discussion

In this section, details of overall results are discussed. The results obtained from classification, clustering and association rule mining are provided in this section.

4.1. Classification Results

Three classification algorithms, ID3, Naïve Bayes and J48, are applied to the suicide bombing attack dataset and the obtained results are as follows. For each technique, the confusion matrix is obtained and the accuracy is compared.

First, we use ID3 to categorize and predict the sensitivity of the location where chances of suicide blast occurrence are high. This will help us in estimating the categories of location for the blast under similar conditions. The Table 1 describes the result where the number of correctly classified records for highly sensitive areas for blasts is 270, which is a high rate of targeted locations in suicide bombing attacks, while 53 attacks are correctly classified in the low category. This implies that suicide attackers mostly target the sensitivity of the locations. The overall accuracy of the ID3 algorithm is 73.8%.

Table 1. Confusion matrix obtained for ID3.

a	b	c	Classified as
270	23	17	a =High
58	56	11	b=Medium
18	7	53	c=Low

The Table 2 describes the confusion matrix of using Naïve Bayes. The result shows that 254 attacks are correctly classified for the highly sensitive areas in suicide bombing attacks, while 48 attacks are correctly classified in the low category. The overall accuracy of Naïve Bayes is 73.2%.

Table 2. Confusion matrix obtained for Naïve Bayes.

a	b	C	Classified as
254	33	23	a=High
41	74	10	b=Medium
17	13	48	c=Low

The Table 3 shows the confusion matrix of using J48. The result shows that 261 attacks are correctly classified for the highly sensitive areas in suicide bombing attacks, while 53 attacks are correctly classified in the low category. The overall accuracy of J48 is 75.4%.

Table 3. Confusion matrix obtained for J48.

a	b	c	Classified as
261	27	22	a=High
41	73	11	b=Medium
14	11	53	c=Low

To interpret the analysis of matrices correctly in evaluating of the classification techniques, we measure the following evaluation metrics: precision, recall, F-measure, MCC and accuracy, as shown in the Table 4. These metrics are most widely used for evaluating the classification performance. These classification methods perform reasonably well for the prediction of possible suicide attack locations. The J48 algorithm shows the highest precision, recall, F-measure, MCC and accuracy values over ID3 and Naïve Bayes.

Table 4. Accuracy values obtained for ID3, Naïve Bayes and J48.

Algorithms	Precision	Recall	F-Measure	MCC	Accuracy
J48	0.753	0.754	0.753	0.559	75.4%
ID3	0.730	0.739	0.728	0.509	73.8%
Naïve Bayes	0.732	0.733	0.733	0.521	73.2%

4.2. Results of Cluster Analysis

To identify similar data, we apply K-means to the dataset. We vary the value for K as 2, 3, 4 and 5 in order to obtain an adequate number of clusters with diverse values of centroids. The result is observed that at K=5 information in each cluster is varied from each other, i.e., each cluster centroid value is distinct from others.

Table 5 shows the K-means results. We observe all the suicide bomb attacks blast on working days in the city of Peshawar, target military organizations and location sensitivity is high, residing in Cluster 0. All the attacks blast on working days in the city of Quetta, where the target peoples are religious, and location sensitivity is medium, residing in Cluster 4. In this analysis, we can see that suicide bombers mostly attacked on working days and targeted highly sensitive areas.

Table 5. K-means clustering results for K = 5.

	Clusters #					
	Full Data	**Cluster0**	**Cluster1**	**Cluster2**	**Cluster3**	**Cluster4**
The Number of Attacks	513	118	79	147	67	102
Blast Day Type	Working Day	Working Day	Working Day	Working Day	Working Day	Working Day
City	Peshawar	Quetta	Peshawar	Rawalpindi	Peshawar	Quetta
Location Category	Police Checkpoint	Market	Religious Places	Police Checkpoint	Police Checkpoint	Religious Places
Target Type	Military	Civilian	Religious	Police	Military	Religious
Location Sensitivity	High	Low	Medium	High	High	Medium
Killed Max	14.924	15.2712	26.2405	8.0136	10.6418	18.5294

4.3. Frequent Pattern Mining Result

We apply the Apriori algorithm to Location Category for pattern identification of attacks on different categories of locations. The attribute Location Category is fixed at the right-hand side of the association rules, while the rest of the attributes are set as default, stated at the left-hand side. The Table 6 describes the sample list of strong and interesting attack patterns generated using the Apriori algorithm showing the support (S), confidence (C) and lift (L) values. The minimum support value for frequent pattern mining is set to 2.

Table 6. Sample strong and interesting frequent patterns mined from the suicide bombing attack dataset.

	Frequent Patterns Mined	**S**	**C**	**L**
1	{Blast.Day.Type='Working Day', Location Sensitivity = High, Target.Type.=foreigner} => {Location Category=Foreign Ambessy}	0.009	1	72
2	{Blast.Day.Type='Working Day',Target.Type.=Media, Location Sensitivity =Medium} => {Location Category=Office Building}	0.008	1	65.429
3	{ Blast.Day.Type='Working Day', Target.Type='Shia sect', Location Sensitivity =High} => {Location Category=Religious Places}	0.008	1	63.923
4	{Blast.Day.Type='Working Day', Target.Type=Religious, Location Sensitivity=Medium} => {Location Category=Religious Places}	0.007	1	57.6
5	{Blast.Day.Type.='Working Day',Target.Type.='Civilian & Police'} => {Location Category=Police Organization}	0.006	1	54.286

As the frequent patterns are identified in the Table 5, the patterns of suicide bombing attack activities are identified. For example, Rule #2 shows that if the Blast Day Type is 'Working Day', Target Type is 'Media' and Location Sensitivity is 'Medium', then there is a high probability of a suicide attack on an 'office Building' on the basis of support 0.008, confidence 1 and lift 65.429, respectively. The result shows the effectiveness of the Apriori algorithm in order to be able to detect newer suicide attacks and unknown patterns in the future.

4.4. Results Validation

The results obtained through the classification algorithm are helpful in predicting the sensitivity of the location where the likelihood of suicide blast occurrence is high. Furthermore, the results acquired using association rule mining support those generated from cluster analysis (examples described in Section 4.2. Cluster Analysis Results and Section 4.3. Apriori Algorithm Results). Cluster analysis is conducted to identify the likelihood of potential suicide attack actions occurring at a specific area, whereas association rule mining identifies common suicide attack patterns of as well as the strength of the link between two events. The findings of both analyses are accurate, implying that similar suicide attack activities and patterns are likely to occur.

5. Conclusion and Future Work

We conclude our research by analyzing the extracted patterns of suicide bomb attacks using data mining strategies based on previous records of a country. We have successfully identified the patterns of suicide attacks that occurred in the past and have also predicted the next possible attack location and expected blast intensity. Using a clustering algorithm, *K*-means, an association rule mining strategy, Apriori, and multiple classification algorithms, Naïve Bayes, ID3 and J48, we have analyzed the relations and corresponding risks. This study has found various similar suicide attack characteristics and has predicted the location of upcoming attacks along with risk intensity pattern identification. Three data mining classification techniques have been applied, i.e., Naïve Bayes, ID3 and J48. The attained accuracy is 73.2%, 73.8% and 75.4%, respectively. That is, J48 provides more accurate results than ID3 and Naïve Bayes. This work can be a significant step toward reducing the world's rising rate of suicide attacks. The findings of this study will assist concerned analysts in cracking and investigating suicide attack cases with relatively little effort.

The data used in this study are limited to Pakistan, but, in the future, we plan to extend the functionalities to predict and identify suicide attacks at a global level through inspiration from recent work of big data Spark ML and big deep learning models [45–53].

Furthermore, the bomb attacks can also be identified through deep learning techniques inspired by recentstudies [54–62].

Author Contributions: Conceptualization, M.K.; Data curation, A.M.Z.; Formal analysis, H.N. and A.Y.; Funding acquisition, H.N.; Investigation, F.F. and A.M.Z.; Methodology, F.F. and M.J.A.; Supervision, M.T.H.; Writing—original draft, F.F. and M.K.; Writing—review and editing, M.T.H., M.J.A. and A.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: This dataset is available online and anyone can use it from the Kaggle website: Pakistan Suicide Bombing Attacks. Available online: https://www.kaggle.com/zusmani/pakistansuicideattacks/notebooks (accessed on 1 December 2017) [16].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Javaid, D. Genesis and effects of religious extremism in Pakistan. *Int. J. Bus. Soc. Sci.* **2011**, *2*, 282–288.
2. Murray, R.M.; Conroy, E.; Connolly, M.; Stokes, D.; Frazer, K.; Kroll, T. Scoping Review: Suicide Specific Intervention Programmes for People Experiencing Homelessness. *Int. J. Environ. Res. Public Health* **2021**, *18*, 6729. [CrossRef]
3. Allen, C. A Critical Analysis of Britain's Living, Dead and Zombie Multiculturalism: From 7/7 to the London 2012 Olympic Games. *Soc. Sci.* **2015**, *4*, 18–33. [CrossRef]
4. Merari, A. *Social, Organizational and Psychological Factors in Suicide Terrorism*; Routledge: London, UK, 2005.
5. Reifels, L.; Morgan, A.; Too, L.S.; Schlichthorst, M.; Williamson, M.; Jordan, H. What Works in Community-Led Suicide Prevention: Perspectives of Wesley LifeForce Network Coordinators. *Int. J. Environ. Res. Public Health* **2021**, *18*, 6084. [CrossRef] [PubMed]
6. Rehma, A.A.; Awan, M.J.; Butt, I. Comparison and Evaluation of Information Retrieval Models. *VFAST Trans. Softw. Eng.* **2018**, *6*, 7–14.
7. Alam, T.M.; Awan, M.J. Domain analysis of information extraction techniques. *Int. J. Multidiscip. Sci. Eng.* **2018**, *9*, 1–9.
8. Mroszczyk, J. To die or to kill? An analysis of suicide attack lethality. *Terror. Political Violence* **2019**, *31*, 346–366. [CrossRef]
9. Poland, J.M. Suicide bombers: A global problem. *Humboldt J. Soc. Relat.* **2003**, *27*, 100–135.
10. Gupta, M.; Jain, R.; Arora, S.; Gupta, A.; Awan, M.J.; Chaudhary, G.; Nobanee, H. AI-enabled COVID-19 outbreak analysis and prediction: Indian states vs. union territories. *Comput. Mater. Contin.* **2021**, *67*, 933–950. [CrossRef]
11. Nagi, A.T.; Awan, M.J.; Javed, R.; Ayesha, N. A Comparison of Two-Stage Classifier Algorithm with Ensemble Techniques on Detection of Diabetic Retinopathy. In Proceedings of the 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA), Riyadh, Saudi Arabia, 6–7 April 2021; pp. 212–215.
12. Anam, M.; Ponnusamy, V.A.; Hussain, M.; Nadeem, M.W.; Javed, M.; Gou, H.G.; Qadeer, S. Osteoporosis Prediction for Trabecular Bone using Machine Learning: A Review. *Comput. Mater. Contin.* **2021**, *67*, 89–105. [CrossRef]
13. Awan, M.J.; Yasin, A.; Nobanee, H.; Ali, A.A.; Shahzad, Z.; Nabeel, M.; Zain, A.M.; Shahzad, H.M.F. Fake News Data Exploration and Analytics. *Electronics* **2021**, *10*, 2326. [CrossRef]

14. Javed, R.; Saba, T.; Humdullah, S.; Jamail, N.S.M.; Awan, M.J. An Efficient Pattern Recognition Based Method for Drug-Drug Interaction Diagnosis. In Proceedings of the 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA), Riyadh, Saudi Arabia, 6–7 April 2021; pp. 221–226.
15. Malik, Z.U.A.; Zhilong, H.; Ashraf, D. Terrorism: The Biggest Security Challenge to the Integrity of Pakistan. *Orient Res. J. Soc. Sci.* **2019**, *4*, 96–106.
16. Usmani, Z.-u.-h. Pakistan Suicide Bombing Attacks. Available online: https://www.kaggle.com/zusmani/pakistansuicideattacks/notebooks (accessed on 1 December 2017).
17. Hassan, J.; Sarfraz, M.S. Impact of suicide bombings in Pakistan using spatial and temporal analysis. In Proceedings of the Seventh International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2019), Paphos, Cyprus, 18–21 March 2019; p. 111740S.
18. Raoof, R.; Rasool, R.; Sattar, J.; Shami, U.T.; Murtaza, G.; Ashraf, M.; Ali, Q.; Malik, A. Evaluation of disaster risk management structure of pakistan with emphasis on man-made disaster. *PalArch's J. Archaeol. Egypt Egyptol.* **2021**, *18*, 1281–1294.
19. Ali, Y.; Farooq, A.; Alam, T.M.; Farooq, M.S.; Awan, M.J.; Baig, T.I. Detection of Schistosomiasis Factors Using Association Rule Mining. *IEEE Access* **2019**, *7*, 186108–186114. [CrossRef]
20. Yousafzai, A.W.; Khan, S.A.; Bano, S.; Khan, M.M. Exploring the phenomenon of suicidal behaviour (SB): An explanatory, mixed-method study in rural Pakistan. *Int. J. Soc. Psychiatry* **2021**. [CrossRef]
21. Asal, V.H.; Park, H.H.; Rethemeyer, R.K.; Ackerman, G. With friends like these . . . why terrorist organizations ally. *Int. Public Manag. J.* **2016**, *19*, 1–30. [CrossRef]
22. Filote, A.; Potrafke, N.; Ursprung, H. Suicide attacks and religious cleavages. *Public Choice* **2016**, *166*, 3–28. [CrossRef]
23. Cozza, V.; Rubino, M. Detection of Similar Terrorist Events. In Proceedings of the IIR, Hangzhou, China, 31 August–2 September 2014; pp. 28–33.
24. Malik, M.S.A.; Sandholzer, M.; Khan, M.Z.; Akbar, S. Identification of risk factors generating terrorism in Pakistan. *Terror. Political Violence* **2015**, *27*, 537–556. [CrossRef]
25. Verma, D.C.; Gartner, S.S.; Felmlee, D.H.; Braines, D.; Yarlagadda, R. Using AI/ML to predict perpetrators for terrorist incidents. In Proceedings of the Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II, Online. 27 April–8 May 2020; p. 114130G.
26. Diab, S. Optimizing stochastic gradient descent in text classification based on fine-tuning hyper-parameters approach. a case study on automatic classification of global terrorist attacks. *arXiv* **2019**, arXiv:1902.06542.
27. Saeed, L.; Syed, S.H.; Martin, R.P. Historical patterns of terrorism in Pakistan. *Def. Secur. Anal.* **2014**, *30*, 209–229. [CrossRef]
28. Imana, E.Y.; Kirk, D. Random walk in extreme conditions-an agent based simulation of suicide bombing. In Proceedings of the 2009 IEEE Symposium on Intelligent Agents, Nashville, TN, USA, 30 March–2 April 2009; pp. 114–121.
29. Conlon, S.J.; Abrahams, A.S.; Simmons, L.L. Terrorism information extraction from online reports. *J. Comput. Inf. Syst.* **2015**, *55*, 20–28. [CrossRef]
30. Saiya, N.; Scime, A. Explaining religious terrorism: A data-mined analysis. *Confl. Manag. Peace Sci.* **2015**, *32*, 487–512. [CrossRef]
31. Tutun, S.; Akça, M.; Bıyıklı, Ö.; Khasawneh, M.T. An outlier-based intention detection for discovering terrorist strategies. *Procedia Comput. Sci.* **2017**, *114*, 132–138. [CrossRef]
32. Uddin, M.I.; Zada, N.; Aziz, F.; Saeed, Y.; Zeb, A.; Ali Shah, S.A.; Al-Khasawneh, M.A.; Mahmoud, M. Prediction of future terrorist activities using deep neural networks. *Complexity* **2020**, *2020*, 1373087. [CrossRef]
33. JSPM, W.; Tirwa, K. Predictive modeling of terrorist attacks using machine learning. *Int. J. Pure Appl. Math.* **2018**, *119*, 49–61.
34. Huamaní, E.L.; Alicia, A.M.; Roman-Gonzalez, A. Machine learning techniques to visualize and predict terrorist attacks worldwide using the global terrorism database. *Mach. Learn.* **2020**, *11*. [CrossRef]
35. Kirk, D.R. Emergency 101-suicide bombers, crowd formations and blast waves. In Proceedings of the MILCOM 2008-2008 IEEE Military Communications Conference, San Diego, CA, USA, 16–19 November 2008; pp. 1–7.
36. Tayal, D.K.; Jain, A.; Arora, S.; Agarwal, S.; Gupta, T.; Tyagi, N. Crime detection and criminal identification in India using data mining techniques. *AI Soc.* **2015**, *30*, 117–127. [CrossRef]
37. Bird, S.; Klein, E.; Loper, E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*; O'Reilly Media, Inc.: Newton, MA, USA, 2009.
38. Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I.H. The WEKA data mining software: An update. *ACM SIGKDD Explor. Newsl.* **2009**, *11*, 10–18. [CrossRef]
39. Rish, I. An empirical study of the naive Bayes classifier. In Proceedings of the IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, Seattle, WA, USA, 4 August 2001; pp. 41–46.
40. Hssina, B.; Merbouha, A.; Ezzikouri, H.; Erritali, M. A comparative study of decision tree ID3 and C4. 5. *Int. J. Adv. Comput. Sci. Appl.* **2014**, *4*, 13–19.
41. Rajesh, P.; Karthikeyan, M. A comparative study of data mining algorithms for decision tree approaches using weka tool. *Advances in Natural and Applied Sciences* **2017**, *11*, 230–243.
42. Sharma, R.; Alam, M.A.; Rani, A. K-means clustering in spatial data mining using weka interface. *Int. J. Comput. Appl.* **2012**, 26–30.
43. Srivastava, S. Weka: A tool for data preprocessing, classification, ensemble, clustering and association rule mining. *Int. J. Comput. Appl.* **2014**, *88*, 26–29. [CrossRef]

44. Tanna, P.; Ghodasara, Y. Using Apriori with WEKA for frequent pattern mining. *arXiv* **2014**, arXiv:1406.7371.
45. Awan, M.J.; Rahim, M.S.M.; Nobanee, H.; Yasin, A.; Khalaf, O.I.; Ishfaq, U. A big data approach to black friday sales. *Intell. Autom. Soft Comput.* **2021**, *27*, 785–797. [CrossRef]
46. Awan, M.J.; Rahim, M.S.M.; Nobanee, H.; Munawar, A.; Yasin, A.; Zain, A.M. Social Media and Stock Market Prediction: A Big Data Approach. *Comput. Mater. Contin.* **2021**, *67*, 2569–2583. [CrossRef]
47. Awan, M.J.; Khan, R.A.; Nobanee, H.; Yasin, A.; Anwar, S.M.; Naseem, U.; Singh, V.P. A Recommendation Engine for Predicting Movie Ratings Using a Big Data Approach. *Electronics* **2021**, *10*, 1215. [CrossRef]
48. Ahmed, H.M.; Awan, M.J.; Khan, N.S.; Yasin, A.; Shehzad, H.M.F. Sentiment Analysis of Online Food Reviews using Big Data Analytics. *Ilkogr. Online* **2021**, *20*, 827–836.
49. Awan, M.J.; Khan, M.A.; Ansari, Z.K.; Yasin, A.; Shehzad, H.M.F. Fake Profile Recognition using Big Data Analytics in Social Media Platforms. *Int. J. Comput. Appl. Technol.* **2021**, in press.
50. Aftab, M.O.; Awan, M.J.; Khalid, S.; Javed, R.; Shabir, H. Executing Spark BigDL for Leukemia Detection from Microscopic Images using Transfer Learning. In Proceedings of the 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA), Riyadh, Saudi Arabia, 6–7 April 2021; pp. 216–220.
51. Awan, M.J.; Gilani, S.A.H.; Ramzan, H.; Nobanee, H.; Yasin, A.; Zain, A.M.; Javed, R. Cricket Match Analytics Using the Big Data Approach. *Electronics* **2021**, *10*, 2350. [CrossRef]
52. Awan, M.J.; Farooq, U.; Babar, H.M.A.; Yasin, A.; Nobanee, H.; Hussain, M.; Hakeem, O.; Zain, A.M. Real-Time DDoS Attack Detection System Using Big Data Approach. *Sustainability* **2021**, *13*, 10743. [CrossRef]
53. Awan, M.J.; Bilal, M.H.; Yasin, A.; Nobanee, H.; Khan, N.S.; Zain, A.M. Detection of COVID-19 in Chest X-ray Images: A Big Data Enabled Deep Learning Approach. *Int. J. Environ. Res. Public Health* **2021**, *18*, 10147. [CrossRef]
54. Banan, A.; Nasiri, A.; Taheri-Garavand, A. Deep learning-based appearance features extraction for automated carp species identification. *Aquac. Eng.* **2020**, *89*, 102053. [CrossRef]
55. Awan, M.J. Acceleration of Knee MRI Cancellous bone Classification on Google Colaboratory using Convolutional Neural Network. *Int. J. Adv. Trends Comput. Sci. Eng.* **2019**, *8*, 83–88. [CrossRef]
56. Abdullah, A.; Awan, M.; Shehzad, M.; Ashraf, M. Fake News Classification Bimodal using Convolutional Neural Network and Long Short-Term Memory. *Int. J. Emerg. Technol. Learn.* **2020**, *11*, 209–212.
57. Awan, M.J.; Rahim, M.M.S.; Salim, N.; Mohammed, M.A.; Garcia-Zapirain, B.; Abdulkareem, K.H. Efficient Detection of Knee Anterior Cruciate Ligament from Magnetic Resonance Imaging Using Deep Learning Approach. *Diagnostics* **2021**, *11*, 105. [CrossRef] [PubMed]
58. Shamshirband, S.; Rabczuk, T.; Chau, K.-W. A survey of deep learning techniques: Application in wind and solar energy resources. *IEEE Access* **2019**, *7*, 164650–164666. [CrossRef]
59. Awan, M.J.; Raza, A.; Yasin, A.; Shehzad, H.M.F.; Butt, I. The Customized Convolutional Neural Network of Face Emotion Expression Classification. *Ann. Rom. Soc. Cell Biol.* **2021**, *25*, 5296–5304.
60. Mujahid, A.; Awan, M.J.; Yasin, A.; Mohammed, M.A.; Damaševičius, R.; Maskeliūnas, R.; Abdulkareem, K.H. Real-Time Hand Gesture Recognition Based on Deep Learning YOLOv3 Model. *Appl. Sci.* **2021**, *11*, 4164. [CrossRef]
61. Fan, Y.; Xu, K.; Wu, H.; Zheng, Y.; Tao, B. Spatiotemporal modeling for nonlinear distributed thermal processes based on KL decomposition, MLP and LSTM network. *IEEE Access* **2020**, *8*, 25111–25121. [CrossRef]
62. Mubashar, R.; Javed Awan, M.; Ahsan, M.; Yasin, A.; Partab Singh, V. Efficient Residential Load Forecasting using Deep Learning Approach. *Int. J. Comput. Appl. Technol.* **2021**, in press.